Volume 12 Number 3 1984 Nucleic Acids Research

The Bam repeats of the mouse genome belong in several superfamilies the longest of which is over 9 kb in size

Michèle Meunier-Rotival and Giorgio Bernardi

Laboratoire de Génétique Moléculaire, Institut Jacques Monod, Tour 43, 2, Place Jussieu, 75005 Paris, France

Received 29 November 1983; Accepted 12 January 1984

ABSTRACT

Mouse DNA contains two equally abundant, homologous subfamilies of MspI 3.6 and 5 kb repeated fragments. The first subfamily corresponds to the previously described (1) Bam 4 kb repeats, the second one to Bam repeated fragments of higher molecular weight. These subfamilies account for the vast majority of long Bam repeats and are linked with contiguous short Bam 0.5 kb repeats. A minority of these composite Bam repeats extend, on the 0.5 kb side, into R repeats. In turn, a fraction of the composite Bam/R repeats extend further, for at least 3 kb, into other repeated sequences contiguous to the R repeats. The long Bam repeats belong, therefore, in at least three superfamilies of repeats, the longest one being over 9 kb in size. Some general properties of these superfamilies are discussed.

INTRODUCTION

Previous work (1) on the long interspersed repeated sequences of the mouse genome showed a) that the EcoRI 1.3 kb repeats (2-6) are contained in long Bam repeats, 4 kb in size; b) that another fraction of the long Bam repeats is present in fragments larger than 4 kb in size; c) that the Bam 4 kb repeats are linked with short Bam 0.5 kb repeats; both Bam repeats are transcribed (7).

In the present investigations, we have obtained a more complete picture of these composite Bam repeats and of other adjacent repeats, as they exist in the mouse genome. First of all, we have been able to show that the long Bam repeats belong in two equally abundant MspI subfamilies, 3.6 and 5 kb in size, which hybridize practically the totality of 4 kb probes. Both subfamilies are contiguous to MspI 0.6 kb repeats which contain the Bam 0.5 kb repeats.

Second, following a very recent report (8) showing that R repeats (9) are linked to Bam 0.5 kb repeats located near immunoglobulin light chain genes, we have studied the general problem of this linkage.
We have shown that the majority of R repeats are linked with Bam 0.5 kb repeats belonging in the Bam 4 - 0.5 kb composite repeats. Furthermore,

a number of these Bam/R composite repeats extend further, for at least 3 kb, into other repeated sequences contiguous to the R repeats.

MATERIALS AND METHODS

DNAs from BALB/c mouse liver or thymus were prepared as described (10). Specific probes for the Bam repeats were plasmids pMRB1-1, pMRB5 and pMRE1 (1; see Fig. 1A). Cloned EcoRI-Bam 0.28 kb and Bam-EcoRI 2.5 kb fragments from pMRB1-2 (pMRB1-21 and -22) were also used as probes (Fig. 1A). The R family probe was a R1/R2-containing subclone (9) obtained from Prof. H.G. Zachau. DNA fragments were transferred from agarose gels to nitrocellulose filters (BA83, Schleicher and Schüll, Dassel, W. Germany) by blotting. Radioactive labeling of DNA, pre-hybridization and hybridization were done as described (11). Isolation and purification of the cloned fragments, 5' or 3' end labelling and sequence analysis were done according to Maxam and Gilbert (12).

RESULTS

The long Bam repeats comprise two sub-families

When mouse DNA is digested with MspI, an enzyme cleaving both CCGG and CmCGG, gel-electrophoresed and stained with ethidium bromide, two bands, 5 and 3.6 kb, appear over the background smear. These bands do not appear after degradation with Hpa II, an enzyme cleaving CCGG and mCCGG, but not CmCGG.

The 5 and 3.6 kb fragments hybridize both the EcoRI 1.3 kb and the Bam 4 kb probes (Fig. 2 A-B; lanes 4), but not the Bam 0.5 kb probe (Fig. 2C; lane 4). In contrast, the same probes only produced a very high molecular weight hybridization smear on Hpa II digests (Fig. 2 A-C; lanes 1); 5 and 3.6 kb bands can, however, be detected in Hpa II digests after very long exposure times (not shown), indicating that the CCGG sites delimiting the MspI fragments are very heavily, but not completely, methylated at the inner C.

As a preliminary to the study of the MspI fragments in genomic DNA, Hpa II sites were mapped on three cloned, (unmethylated), Bam 4 kb fragments. This showed the presence of one, two or three MspI sites; sites 2 and 5 were found in two fragments (Fig. 1A); site 5 is almost coincident with the Bam site forming the right end of the fragments (Fig. 6), as presented in Fig. 1A (where orientation is reversed relative to that shown in ref. 1).

A study of double and triple digests of genomic DNA with MspI and

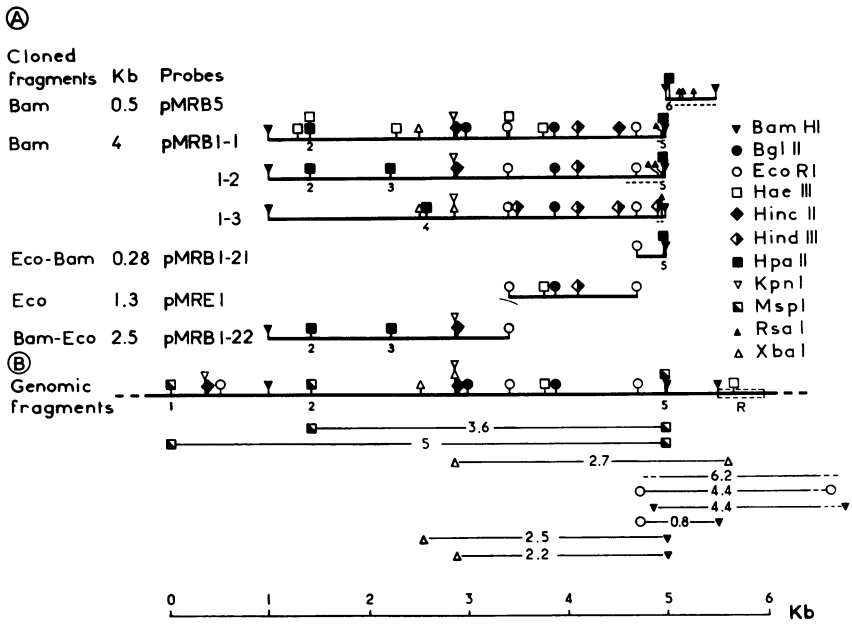

Fig. 1 : A. Restriction maps of cloned fragments from Bam repeats. HpaII site 5 was also found in C_κ, $C_\lambda A$ and $C_\lambda B$ (8). The right HincII site of pMRB1-1 and the right HindIII site of pMRB1-3 were found in the sequence of pMD135 (13). The common HindIII site of pMRB1-1,2,3 was found in three MIF-1 clones (32). The common BglII site of pMRB1-1,2,3 at 1100 bp from the right Bam site corresponds to the orientation II of Gebhard and Zachau (19). Sequenced regions (see Fig. 6) are underlined.

B. Restriction map of genomic fragments. Only frequent sites are indicated. The box indicates the location of the R repeats linked with Bam repeats. Map orientation is reversed relative to that previously shown in order to match the sequences of Fig. 6.

four restriction enzymes previously used (1) allowed us to construct a restriction map of MspI fragments (Fig. 1B). Some salient points are the following : (i) Both MspI fragments contain the two EcoRI sites delimiting the 1.3 kb fragments (Fig. 3 A-B; lanes 6). The other fragments hybridizing the 4 kb probes (Fig. 3B; lane 6) correspond to the EcoRI-MspI 2.1 kb, EcoRI 2.9 kb, EcoRI-MspI 3.3 kb fragments derived from the region at the left of the EcoRI 1.3 kb fragments (see Fig. 1B). (ii) The two genomic MspI fragments, 5 and 3.6 kb, are not cut by Bam (Fig. 2B; lane 3). This indicates that the 3.6 kb fragments are contained in the Bam 4 kb fragments (from which MspI releases the expected MspI-Bam 0.4 kb fragments; not shown), and that the 5 kb fragments either lack, or are not cut at, the Bam site

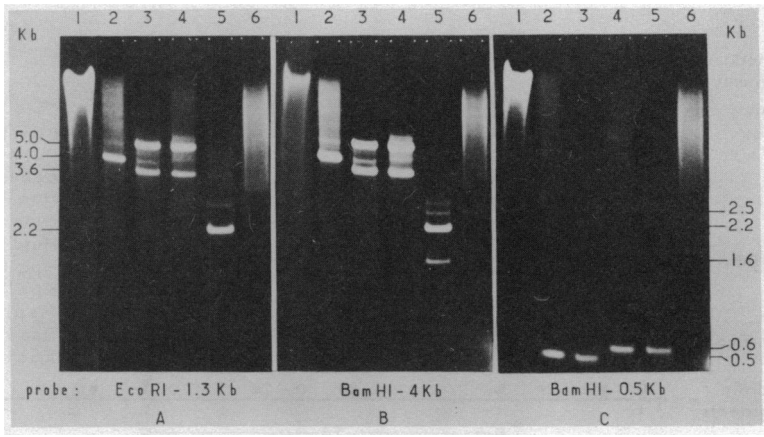

Fig. 2 : Autoradiograms obtained after hybridizing the labelled probes indicated in the figure on the restriction digests of mouse liver DNA (1 : 5 µg; 2-6 : 2.5 µg) after transfer from 1 % agarose gels to nitrocellulose.
1, HpaII; 2, BamHI; 3, BamHI + MspI; 4, MspI; 5, MspI + HincII; 6, HincII.

forming the left end of the 4 kb fragments. (iii) The high molecular weight smear obtained by hybridizing the Bam 4 kb probe on Bam digests (Fig. 2B; lane 2) is practically eliminated when MspI is added to this enzyme (Fig. 2B; lane 3). This indicates that the repeated sequences from the Bam 4 kb region responsible for the smear are contained in the MspI 5 kb fragments. Likewise,

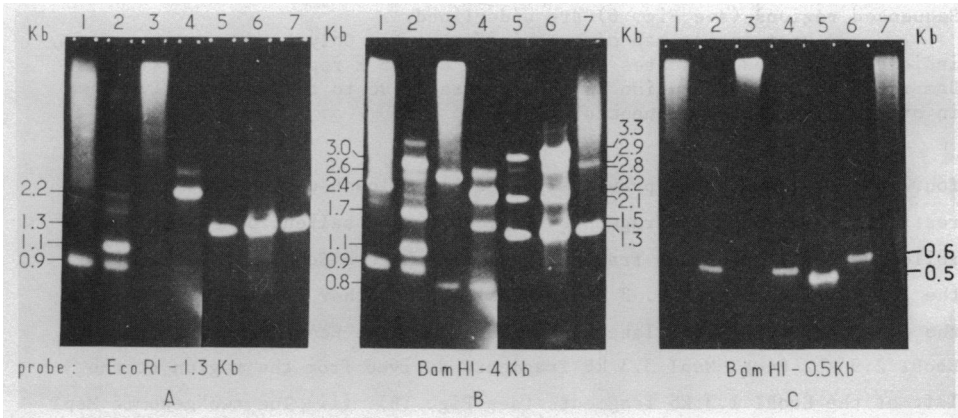

Fig. 3 : Autoradiograms obtained after hybridizing the labelled probes indicated in the figure on the restriction digests of mouse liver DNA (1 µg) after transfer from 1 % (A-B) or 1.5 % (C) agarose gels to nitrocellulose.
1, BglII; 2, BglII + MspI; 3, KpnI; 4, KpnI + MspI; 5, BamHI + EcoRI + MspI; 6, EcoRI + MspI; 7, EcoRI.

the high molecular weight smears obtained by hybridizing the Bam 4 kb probe on HincII (Fig. 2B; lane 6) BglII, KpnI, and EcoRI digests (Fig. 3B;lanes 1, 3, 7) disappear when MspI is added (Fig. 2B; lane 5. Fig. 3B; lanes 2,4,6), indicating that the MspI site responsible for this disappearance is to the left of these previously mapped restriction sites (1). Incidentally, these sites had already provided evidence for an extension of Bam 4 kb fragments. Finally, the two MspI sub-families and their methylation patterns were also found in DNAs from other tissues and cell lines of M.musculus and from the liver of M. spretus (not shown).

The short Bam repeats

Hybridization of the Bam 0.5 kb probe on MspI digests produced a 0.6 kb band (Fig. 2C; lane 4) and two faint 5.6 and 4.2 kb bands (also produced by the EcoRI 1.3 kb and the Bam 4 kb probes; Fig. 2 A-B; lanes 4). While one of the MspI sites delimiting the 0.6 kb fragment is likely to be site 5 or 6 detected on cloned Bam fragments (Fig. 1A), the other one is external to the fragment. The high molecular weight smears found with the Bam 0.5 kb probe on HincII (Fig. 2C; lane 6), BglII, KpnI, EcoRI (Fig. 3C; lanes 1,3,7) disappear upon MspI digestion to give rise to the 0.6 kb bands (Fig. 2C; lane 5; Fig. 3C; lanes 2,4,6). The smears are mostly due to fragments delimited to the left by sites present in the long Bam repeats and to the right by sites located in single-copy sequences.

The composite Bam repeats

Previous evidence (1) showed that EcoRI 1.3 kb and Bam 4 and 0.5 kb probes hybridized on a common HaeIII 1.86 kb band. The physical linkage of the long and short Bam repeats is confirmed here by the hybridization of the same three probes on a common XbaI 2.7 kb band (Fig. 4 A-D; see also next section). The corresponding HaeIII and XbaI sites are indicated in Fig. 1B. The HaeIII and XbaI hybridization patterns indicate that the vast majority of long and short Bam repeats are linked in these composite Bam repeats since the hybridization bands correspond to most of the bound radioactivity. Additional evidence for the physical linkage of the Bam repeats comes from the common hybridization of the two Bam probes on MspI 5.6 and 4.2 kb fragments (Fig. 2 B-C; lanes 4); these faint hybridization bands, due to the absence of MspI site 5 in some composite repeats, disappear upon Bam digestion, as expected.

A fraction of composite Bam repeats is linked with R repeats and extends beyond them

When probes for individual, contiguous segments of the composite

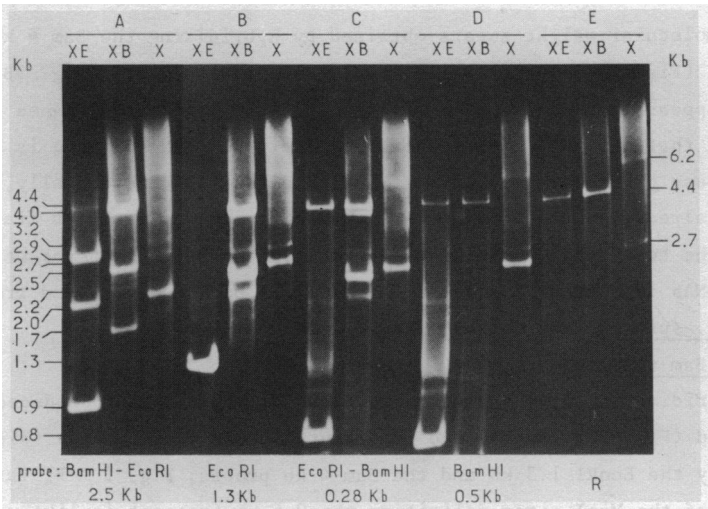

Fig. 4 : Autoradiograms obtained after hybridizing the labelled probes indicated in the figure on XbaI (X), XbaI + BamHI (XB), and XbaI + EcoRI (XE) digests of mouse DNA (5 µg) after transfer from 1 % agarose gels to nitrocellulose.

Bam repeats (Bam-EcoRI 2.5 kb, EcoRI 1.3 kb, EcoRI-Bam 0.28 kb, Bam 0.5 kb) and for R repeats were hybridized on simple, double and triple XbaI (X), XbaI + Bam (XB), XbaI + EcoRI (XE) and XbaI + Bam + EcoRI(XBE) digests, results bearing on the linkage Bam/R and on the right ends of the composite repeats were obtained.

i. All the probes used showed the presence in X digests of 2.7 kb fragments (Fig. 4) delimited by one site located at the left of the EcoRI 1.3 kb fragment (and possibly corresponding to the right XbaI site of pMRB1-3) and one site located in the R repeats (Fig. 1B). The 2.7 kb fragments (which disappear in XE and XB digests, because they contain both EcoRI and Bam sites) provide evidence for a physical linkage of R and Bam composite repeats. Even if the left border of the 2.7 kb fragments is located at only about 0.5 kb to the left of the EcoRI 1.3 kb fragments, this linkage must involve whole Bam 4 kb fragments because essentially all Bam 4 and 0.5 kb are linked together (see preceding section).

ii. All the probes used on X digests produce a hybridization smear (Fig. 4) which is stronger with probes derived from the left side of the composite repeats. In addition, a 6.2 kb hybridization band is evident when using R, Bam 0.5 kb and EcoRI-Bam 0.28 kb probes, whereas the inten-

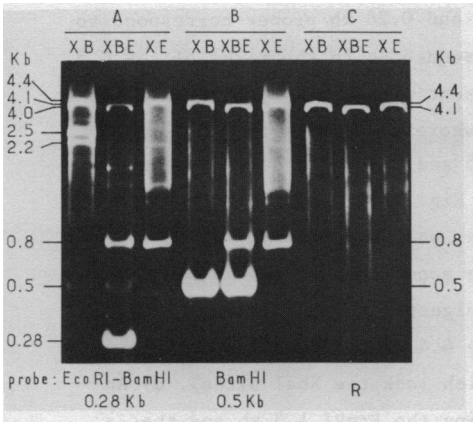

Fig. 5 : Autoradiograms obtained after hybridizing the probes shown on the figure on XbaI + BamHI (XB), XbaI + BamHI + EcoRI (XBE) and XbaI + EcoRI (XE) digests of mouse DNA (5 µg) after transfer from a 2 % agarose gel to nitrocellulose.

sity of the smears obtained with the other probes makes impossible to decide whether the band is present or not in those cases too (Fig. 4 A-B). While the right end of the 6.2 kb fragments must obviously be located in repeated sequences extending R sequences to the right, the left end cannot be positioned on the basis of the results obtained on single X digests (Fig. 2B).

iii. The hybridization banding patterns on XB, XE and XBE digests are essentially due to Bam and EcoRI fragments since the latter digests are identical to those of Figs. 4 and 5, except for very few additional bands and for a decrease of the background smear. a) When using the R probe on XB and XE digests, the most noticeable difference with X digests is that both the high molecular weight smear and the 6.2 kb band disappear to originate a 4.4 kb band; the size of this band decreases to 4.1 kb in XBE digests (Fig. 5). The same 4.4 and 4.1 kb bands are obtained with single Bam and EcoRI digests and with double Bam-EcoRI digests, respectively (not shown). b) When the Bam 0.5 kb probe is used on XB, XE, XBE digests (or on B, E and BE digests), the results concerning the 4.4/4.1kb bands are the same as those just described; in addition, however, the XE or E digests show a strong smear and a strong 0.8 kb band, the XB or B digests show a very strong 0.5 kb band, and the XBE or BE digests show both the 0.8 and the 0.5 kb bands (Fig. 5). c) As far as the 0.8 kb band is concerned, the identical results obtained on XE or E digests with the Bam 0.5 kb and EcoRI-Bam 0.28 kb probes show that the corresponding fragments must be delimited by the right EcoRI site of the 1.3 kb fragment and by the Bam site located at the 0.5 kb/R border. The heavy smears in

XE digests hybridized with both the 0.5 and 0.28 kb probes correspond to XbaI and EcoRI sites located in unique sequences to the right of the 0.5 kb fragments, since these smears are absent with the R probe; these smears are also absent in XB and XBE digests, where the very strong 0.5 and 0.28 probe are 1) the appearance of 2.5 and 2.2 kb bands in XB digests; these are due to fragments indicated in Fig. 1B; (expectedly, the same bands appear with the EcoRI 1.3 kb probe, whereas only the 2.5 kb appears with Bam-EcoRI 2.5 kb probe); and 2) the appearance of a 4 kb band, in addition to the 4.4 kb band in XB or B digests; this indicates that the 0.28 kb probe hybridizes not only on the 4.4 kb fragments extending to the right, but also on Bam 4 kb repeats (which lack the XbaI sites). d) An important result obtained when hybridizing the EcoRI 1.3 kb and the Bam-EcoRI 2.5 kb probes is the absence of the 4.4 band in the XE and XB (or E and B) digests, because this defines the location of the left end of these fragments.

In conclusion, these results confirm that a large fraction of composite Bam repeats extend to the right directly into non-repeated sequences. Another fraction (corresponding to the XbaI 2.7 and 6.2 kb fragments and to 4.4/4.1 kb fragments present in Bam, EcoRI and Bam + EcoRI) not only is contiguous to R repeats, but extends further to the right into other repeated sequences. The high molecular weight smears indicate that yet another fraction ends in non-repeated sequences.

It is of interest to localize the ends of both the 6.2 and 4.4/4.1 kb fragments. The absence of a 4.4 kb band in XE (and E) digests hybridized with the EcoRI 1.3 kb and the Bam-EcoRI 2.5 kb probes (Fig. 4 A-B) indicates that the left end of these fragments coincides with the left end of the EcoRI 0.28 kb fragment. The right end of the 4.4 kb fragment must therefore be located at least 3 kb further downstream of R repeats. In the case of XB (and B) digests, the 4.4 kb fragments are delimited by two Bam sites, 4.4 kb apart from each other and 0.15 kb to the right of the EcoRI sites just discussed. This positioning is the only one accounting for the 4.1 kb fragments in double XBE (and BE) digests.

It should be stressed that the 4.4/4.1 kb fragments lack the Bam sites which delimit most of the 0.5 kb fragments. This raises the question whether the 0.5 kb fragments delimited by such sites are also linked with R repeats. The very strong intensity of the Bam 0.5 kb bands and the very faint smears obtained with R probes on XB and XBE digests showing approximately equal 4.4/4.1 kb bands suggest that only a small minority of these

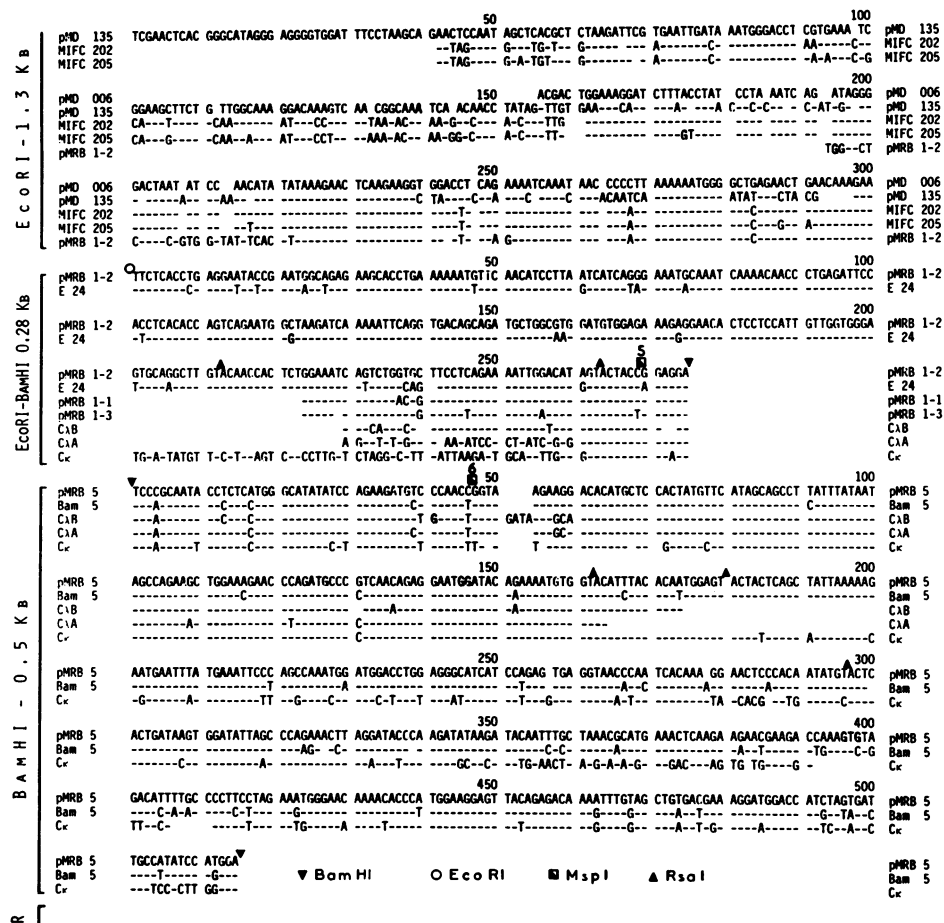

Fig. 6 : Sequence of the rightmost 1 kb stretch from Bam repeats (see Fig. 1A). Our data on pMRB1-1,2,3 and pMRB5 are compared with those on pMD 006 and 135 (13), MIFC 202 and 205 (14), E24 (15), C_κ, $C_\lambda A$, $C_\lambda B$ (8), and Bam 5 (16). The R sequence present in C_κ is not shown. Sequences were aligned so as to maximize homologies. The orientation presented is that of most published data.

Bam 0.5 kb fragments are linked with R repeats.

Finally, the precise positioning of the 6.2 kb fragments is not established yet; if their left ends coincide with that of the 4.4 kb fragments, then repeated sequences would extend by at least 5 kb beyond the R repeats.

The primary structure of the BamHI 0.5 kb fragments and of its flanking sequences

Fig. 6 compares the sequences determined in this work for the right end of the Bam 4 kb fragments and for the Bam 0.5 kb fragments with recent literature data (13-16). Concerning the first segment, the right end of the EcoRI 1.3 kb fragments, three of the published sequences (13, 14) diverged by less than 9 % from each other; a fourth sequence (13) diverged, however, by 36 % from the other ones; our own sequence diverged by about 9 % from the first three over 80 bp. The EcoRI-Bam 0.28 kb fragments appear to be relatively conserved (6-17 % divergence) in five sequences (8, 15 and this work), whereas two others (8) only share a few nucleotides with them.

As far as the Bam 0.5 kb fragment is concerned, our sequence diverges from Bam 5 (16) by 10 %, and from C_κ (8) by 22 %. These differences are more frequent between positions 201 and 515, where our sequence diverges from Bam 5 by 18 % and from C_κ by 44 %. In high-divergence regions, deletions are present and base changes are highly clustered.

Finally, Bam repeats are 43 % in GC over 1 kb, and show an asymmetric distribution of A (and T), which ranges between 33 and 44 % on one strand.

DISCUSSION

The two MspI subfamilies of Bam 4 kb repeats

Hybridization of the EcoRI 1.3 kb or the Bam 4 kb probes on MspI digests of mouse DNA defines two equally abundant sub-families of repeats, (Fig. 2 A-B), 5 and 3.6 kb in size, corresponding to sites 1 and 5, and sites 2 and 5, respectively (Fig. 1B). As far as the size of the repeats defined by MspI digestion is concerned, while one sub-family is at least 5 kb long, the other one is at least 4 kb long, reaching the left end of the Bam 4 kb fragments. The latter subfamily may in fact extend as far as the other one; if it does so, the repeats lack MspI site 1, since the corresponding 1.4 kb fragments (comprised between sites 1 and 2) were not seen. The main interest of finding the two MspI sub-families is that they account for essentially the totality of the hybridization of EcoRI 1.3 kb and Bam 4 kb probes and establish a 1:1 relationship for the EcoRI 1.3 kb and the Bam 4 kb fragments.

The Bam 0.5 kb repeats and their linkage with the Bam 4 kb repeats

A probe for these repeats hybridizes on MspI 0.6 kb fragments (Fig. 2 C), which are delimited by MspI sites 5 or 6 on the left and an external MspI site on the right. MspI site 6 probably is not a very frequent site since it was only found in pMRB5 out of five sequences (Fig. 6). In contrast, MspI site 5 was found in 5 out of 7 sequences available so far (Fig. 6).

The linkage of the Bam 4 and 0.5 kb fragments, originally established when a common 1.86 kb hybridization band was found in HaeIII digests (1), is now confirmed by other common hybridization bands : XbaI 2.7 kb (Fig. 4 A-B) and MspI 5.6 and 4.2 kb (Fig. 2 B-C; lane 4). The orientation of the two families of Bam fragments in the composite Bam repeats (Fig. 2B) is based on the sequencing results of Fig. 6. Our data do not provide any information about the left end of the repeats.

Finally, methylation is very frequent at MspI sites, as shown by their resistance against Hpa II degradation, and might also account for the resistance of some Bam sites (GGATmCC is not cut by Bam; 17).

The linkage of R repeats with the Bam composite repeats

A linkage between R and Bam 0.5 kb repeats was first reported (8) in composite repeats found near the immunoglobulin light chain genes and the conclusion was drawn that such composite repeats did not extend beyond the 0.5 kb repeats and formed a class by themselves (8).

Our results show, however, that the general situation in the mouse genome is different, in that the majority of R repeats are linked not only with the Bam 0.5 kb repeats, but also with the contiguous Bam 4 kb repeats in agreement with some very recent, independent findings on cloned material (18, 19). The opposite is not true, however, as shown by the fact that composite Bam repeats are much more abundant than R repeats and often extend directly into single-copy DNA on the right of the map of Fig. 1B.

Four classes of Bam-R composite repeats have been identified in this work. The first one corresponds to the precisely mapped Xba I 2.7 kb fragments (Fig. 1B). The second one corresponds to the EcoRI and Bam 4.4 kb fragments. These have their left ends at the EcoRI site of the 0.28 kb fragment and 0.15 kb to the right of it, respectively, and their right ends in other repeated sequences at least 3 kb beyond the end of the R repeats. This means that the overall length of these Bam-R repeats is no less than 9 kb. The other two classes are less well defined. They comprise the 6.2 kb

fragments and the high molecular weight fragments seen in XbaI digests. It is possible, but not proven, that the right ends of the 6.2 kb fragments, which still are in repeated sequences, are beyond those of the 4.4 kb fragments, whereas those of the long fragments responsible for the high-molecular weight smear are in single-copy sequences further to the right.

General conclusions

In the present paper, we have called subfamilies the sets of repeated sequences which cross-hybridize with each other or with a common probe (e.g. the MspI 5 and 3.6 kb subfamilies), and families the sets which do not (e.g.the Bam 4 and 0.5 kb families). It is now clear that such families, may be linked together into composite families of repeats, which we will call superfamilies. The first such superfamily which was described was the Bam 4/0.5 kb superfamily (1), the second one the Bam 0.5 kb/R superfamily (8). Obviously, these superfamilies may exist as such or belong in larger superfamilies. For instance, the longest superfamily seen in the present work comprises at least four families of repeats, the Bam 4 kb and 0.5 kb families, the R family, and the family(ies) located to the right of R repeats. While members of a single family may exist as isolated repeats, most seem to belong in one or more superfamilies. We will now consider some general properties of these superfamilies, namely properties concerning most of the repeats involved.

The first general property, well-documented by all investigations done so far, is the polymorphism of the repeats which is related to their sequence divergence. The primary structure results presented here indicate an average divergence of 10-20 %, but higher values may be associated with particular sets of repeats like the Bam 0.5 kb/R repeats near immunoglobulin light chain genes. Some regions of the repeats are definitely more conserved than other ones, as witnessed by the different frequencies of restriction sites in the repeats.

A second general property is conservation of family order and family size in the superfamily. For example, the order of the two Bam families or that of the Bam families and the R family is always the same. Likewise, the size of the Bam 0.5 kb fragments, as present in the Bam or in the Bam/R superfamilies, is constant; in other words, truncation within a family does not appear to be a common phenomenon. It should be stressed that this phenomenon is difficult to judge at the restriction fragment level because both the polymorphism of repeats and their truncation yield

fragments forming a smear. In fact, truncation is difficult to assess even at the sequence level. For instance, a local area of higher divergence, such as that exhibited by the right end of the Bam 0.5 kb fragment of C_K (Fig. 6) could easily be mistaken for a truncation, if sequence analysis was not showing a high homology in the flanking sequences and a continuity with R repeats. In contrast, a real truncation appears to exist at the left end of the C_K repeat (Fig. 6), but this truncation, at about 15 bp from the end of the Bam 0.5 kb segment can hardly be considered as occurring within a family and rather is an example of the fact that the ends of repeat families, as defined by certain restriction sites, are indeed close to the real ends of repeat families.

Another general property is the mobility of repeat families and superfamilies. It is obvious that repeat amplification and spreading through the genome has been operational in generating the thousands of interspersed repeat copies present in the mouse genome. This raises the problem of the molecular mechanisms involved in such processes. Since evidence for the transcription of the Bam and R repeats is available (7, 16, 19), it has been strongly suggested (18) that amplification and spreading were achieved via retrotranscription processes starting at the right ends of R sequences. The evidence quoted in support of this suggestion were the poly A tracts at those ends, and the greater abundance of the sequences forming the right ends of the composite repeats relative to those forming the left ends (18). Other findings along this line were the higher copy number of R repeats (100,000 copies; 9) relative to the short Bam repeats (25,000-50,000; 16) and the existence of sequences truncated after the Bam 0.5 kb repeats (8).

Our observations, which do not concern just a few cloned repeats, but provide an overall picture of the repeats as they exist in the genome, do not support the general validity of the facts quoted in favor of a retrotranscription mechanism. Indeed, our results show : 1) that about half of the long Bam repeats are at least 5 kb in size, the other half being at least 4 kb in size, and possibly extending as far as the former; in other words, we have no evidence of frequent truncations cutting in the long Bam repeats, which in fact are stoichiometric with the internal EcoRI 1.3 kb repeats; 2) that the linkage of the short and long Bam repeats involves the vast majority of the repeats; 3) that the linkage of Bam 4/0.5 kb repeats and R repeats is unbalanced, but by a serious shortage of R repeats (the majority of which is linked) and not viceversa; 4) that the poly-A end of R repeats are not the ends of the superfamily in the

fraction of Bam/R composites extending further to the right by at least 3 kb; 5) that the excess copy number of R repeats over Bam repeats, which rests on an estimate of R repeats (100,000; 9; reassessed as 10,000 by other authors; 8), is contradicted by our results indicating that, whatever their actual number, R repeats are much less abundant than Bam repeats; 6) that the explanations given for the Bam/R composite repeats obviously do not apply to the majority of the Bam superfamily which is not linked with R repeats.

Under these circumstances, other evidences for retrotranscription should be looked for, and alternative mechanisms for repeat spreading, in particular transposition, should not be neglected. In this connection, it should be mentioned here that the existence of long terminal repeats delimiting a 7 kb repeat has been claimed by some authors (20) and denied by others (18), and that sequence homologies between intracisternal A particle genes (21) and the repeats have been claimed (20, 22).

A fourth general property of the long repeats under consideration is their non-uniform genomic distribution. Indeed, the Bam composite repeats are almost only found in the two light major components (1) which represent about 2/3 of the mouse genome (10). These findings indicate that, whatever the basic mechanism underlying their mobility, the integration of the repeats is targeted towards particular genome domains. The base composition of Bam repeats is 39 % for the 4 kb fragment (1) and 43 % for the 1 kb encompassing the 0.5 kb fragment (this work); these values are close to the base composition (38 and 40 % GC) of the major components of the mouse genome which harbor them (10). This suggests that the targeted integration of the repeats is directed by sequence homology. Interestingly, a genomic distribution similar to that exhibited by Bam repeats in the mouse genome, was also found for KpnI sequences in the human genome (7). This indicates that these long repeated sequences (or LINES, 23) are evolutionarily conserved not only in primary structure and transcription properties (15,24,25), but also in genomic distribution. This is also true for the short repeated sequences (or SINES, 23) of the Alu family, as found in the human, mouse and chicken genome; in this case, the repeats predominate, however, in the heavy components (7, 26). A final point concerning the genomic distribution of the Bam/R repeats is that it does not seem to match that found for a number of genes (paper in preparation) which are present in all DNA components and are possibly even more abundant in the heavy components. This different distribution suggests that

there is no obvious interpretation for the finding of some Bam/R repeats in the neighborhood of a number of genes (27-31).

ACKNOWLEDGEMENTS

We thank H.G. Zachau for the R probe and for communicating results prior to publication; C. Gaillard for her help with sequencing experiments; D. Bucchini and A. Minty for supplying mouse cell lines and M.spretus DNA.

Abbreviations : kb : kilobase ; bp : base pair; Bam : BamHI.

REFERENCES

1. Meunier-Rotival, M., Soriano, P., Cuny, G., Strauss, F. and Bernardi, G. (1982) Prcc. Natl. Acad. Sci. USA 79, 355-359.
2. Hörz, W., Hess, I. and Zachau, H.G. (1974) Eur. J. Biochem. 45, 501-512.
3. Manuelidis, L. (1980) Nucleic Acids Res. 8, 3247-3258.
4. Heller, R. and Arnheim, N. (1980) Nucleic Acids Res. 8, 5031-5042.
5. Cheng, S.M. and Schildkraut, C.L. (1980) Nucleic Acids Res. 8, 4075-4090.
6. Brown, S.D.M. and Dover, G. (1981) J. Mol. Biol. 150, 441-466.
7. Soriano, P., Meunier-Rotival, M. and Bernardi, G. (1983) Proc. Natl. Acad. Sci. USA 80, 1816-1820.
8. Wilson, R. and Storb, U. (1983) Nucleic Acids Res. 11, 1803-1817.
9. Gebhard, W., Meitinger, T., Höchtl, J. and Zachau, H.G. (1982) J. Mol. Biol. 157., 453-471.
10. Cuny, G., Soriano, P., Macaya, G. and Bernardi, G. (1981) Eur. J. Biochem. 115, 227-233.
11. Meunier-Rotival, M., Cortadas, J., Macaya, G. and Bernardi, G. (1979) Nucleic Acids Res. 6, 2109-2133.
12. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
13. Dubnick, M., Chou, J., Petes, T.D. and Farber, R.A. (1983) J. Mol. Evol. 19, 115-121.
14. Brown, S.D.M. (1983) Gene, in press.
15. Singer, M.F., Thayer, R.E., Grimaldi, G., Lerman, M.I. and Fanning, T.G. (1983) Nucleic Acids Res. 11, 5739-5745.
16. Fanning, T.G. (1982) Nucleic Acids Res. 10, 5003-5013.
17. Mc Clelland, M. (1981) Nucleic Acids Res. 9, 5859-5866.
18. Fanning, T.G. (1983) Nucleic Acids Res. 11, 5073-5091.
19. Gebhard, W. and Zachau, H.G. (1983) J. Mol. Biol. 170, 255-270.
20. Georgiev, G.P., Kramerov, A.P., Ryskov, A.P., Skryabin, K.G. and Lukanidin, E.M. (1982) Cold Spring Harbor Symp. Quant. Biol. 57, 1109-1121.
21. Lueders, K.K. and Kuff, E.L. (1980) Proc. Natl. Acad. Sci. USA. 77, 3571-3575.
22. Brown, A. and Huang, R.C.C. (1982) Proc. Natl. Acad. Sci. USA. 79, 6123-6127.
23. Singer, M.F. (1982) Int. Rev. Cytol. 76, 67-112.
24. Lerman, M.I., Thayer, R.E. and Singer, M.F. (1983) Proc. Natl. Acad. Sci. USA. 80, 3966-3970.
25. Shafit-Zagardo, B., Brown, F.L., Zavodny, P.J. and Maio J.J. (1983) Nature, 304, 277-280.
26. Olofsson, B. and Bernardi, G. (1983) Biochim. Biophys. Acta, 740, 339-341.

27. Haigwood, N.L., Jahn, C.L., Hutchinson III, C.A. and Edgell, M.H. (1981) Nucleic Acids Res. 9, 1133-1150.
28. Lueders, K.K. and Paterson, B.M. (1982) Nucleic Acids Res. 10, 7715-7729.
29. Soriano, P., Szabo, P. and Bernardi, G. (1982) EMBO J. 1, 579-583.
30. Minty, A.J., Alonso, S., Guénet, J.L. and Buckingham A.M. (1983) J. Mol. Biol. 167, 77-101.
31. Burton, F.H., Voliva, C.F., Edgell, M.H. and Hutchinson III, C.A. (1983) DNA 2, 82.
32. Brown, S.D.M. and Piechaczyk, M. (1983) J. Mol. Biol. 165, 249-256.